

For publication

## **Advancing Art Libraries and Curated Web Archives: A Community Survey**



November 14, 2018

---

**Internet Archive**

**New York Art Resources Consortium**

Persistent URL: <https://archive.org/details/AALCWACommunitySurvey>



*Institute of Museum and Library Services (IMLS) National Leadership Grant [LG-88-18-0069-18](#).*



*Published by Archive-It ([archive-it.org](https://archive-it.org)), a web archiving service of the Internet Archive ([archive.org](https://archive.org)).*

*This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.*

## Table of contents

<b>INTRODUCTION</b>	<b>3</b>
<b>SUMMARY</b>	<b>3</b>
<b>RESULTS</b>	<b>5</b>
Institution type	5
Respondent role and experience	6
Born-digital collecting	8
Web archiving familiarity and experience	8
Current practices and interests	10
Barriers to web archiving	15
Web archiving training	16
Interests in collaboration	18
Budgets	19
Descriptive practice	20
<b>ANALYSIS</b>	<b>22</b>
Community	22
Collecting	23
Obstacles and opportunities	24
<b>REFERENCES</b>	<b>26</b>

## INTRODUCTION

The Internet Archive and the New York Art Resources Consortium (NYARC) conducted a survey of the art and museum libraries community in preparation for a National Forum, to consist of a one-day symposium and half-day of technology workshops on the topic of this community's creation and/or use of curated web archives. The survey was open from September 10 to October 6, 2018, and was distributed among the official email lists of the Art Libraries Society of North America (ARLIS/NA), Museum Computer Network (MCN), and Museum Section of the Society of American Archivists (SAA). Respondents were asked for information about their professional roles and organizational typologies, knowledge of and direct experience with web archiving, current collecting practices, and their obstacles to and/or aspirations for web archiving moving forward. The results of this survey, summarized below, are being analyzed actively towards programming a productive National Forum among key stakeholders within and allied with the community of practice, to be held at the San Francisco Museum of Modern Art (SFMOMA) in February, 2019.

This summary report complements a review of published and open literature about the needs for, experiences of, and challenges facing curated web archives in institutions missioned to curate, preserve, and provide access to the art historical record, including born-digital primary source material. The literature review (Internet Archive & New York Art Resources Consortium, 2018) summarizes a well researched and articulated desire for engagement in web archiving among art and museum libraries since at least 2013, and the concurrent efforts to pilot model and establish permanent programs or roles in these libraries or their consortia. This community survey enhances the findings of the literature review by documenting the perspectives of more art and museum library professionals than has been collected prior, providing new insight into the state of web archiving among their institutions, needs and desires of web archives, and practical obstacles to collection development or sustainability.

## SUMMARY

This survey's community of respondents represent diverse roles and responsibilities among art and museum libraries, but share many common perspectives on the needs for web archives. These common ideas and goals, along with key differentiating factors, are detailed in the *Results* and discussed in the *Analysis* sections of this report. In total, 77 community members responded to the survey, 20 of whom volunteered that they already engage in some form of web archiving actively at their respective institutions. The respondents are largely familiar with and even trained to some degree in web archiving, though a clear minority of their institutions have formal policies for collecting born-digital materials generally and extremely few dedicate specific staff to the tasks of web archiving. A large plurality of current and the majority of aspirant web archivists in this community prioritize collecting that will complement thematic or subject-based strengths, but

a majority of both respondent types most highly prioritize preserving the web presences of their own institutions. They accord with the published literature that web archive value cases, discovery and use options, and curation tools all present real but surmountable challenges compared to strategic collecting considerations. Unlike the literature, they indicate less immediate, pressing concern for or interest in collaborative collection development or access than they do for the requisitioning of sufficient budget and staff to begin and sustain web archiving.

To best represent this community and its primary interests and concerns at a National Forum, the event's program should include:

- Librarians, archivists, and managerial decision makers from a representative cross-section of institution types, most importantly museums.
- In-depth analysis of the financial and human resources appropriated by art and museum libraries to create and sustain web archives.
- Thorough discussion of the opportunities to support the above appropriations through collaboration among peer institutions, particularly in long-term material need areas like storage and preservation, but also collection development decisions, descriptive cooperation, curation tool development and access, and training or educational resources.
- Analysis of the specific impacts of formal collecting policies for born-digital media more generally among art and museum libraries on their opportunities to engage web archiving.
- Training opportunities scaled to the needs of professionals with different levels of prior training and even direct experience with web archiving, to transcend the basic, introductory offerings made prior and to include high level policymaking and advanced web data stewardship modules.

## RESULTS

### Institution type

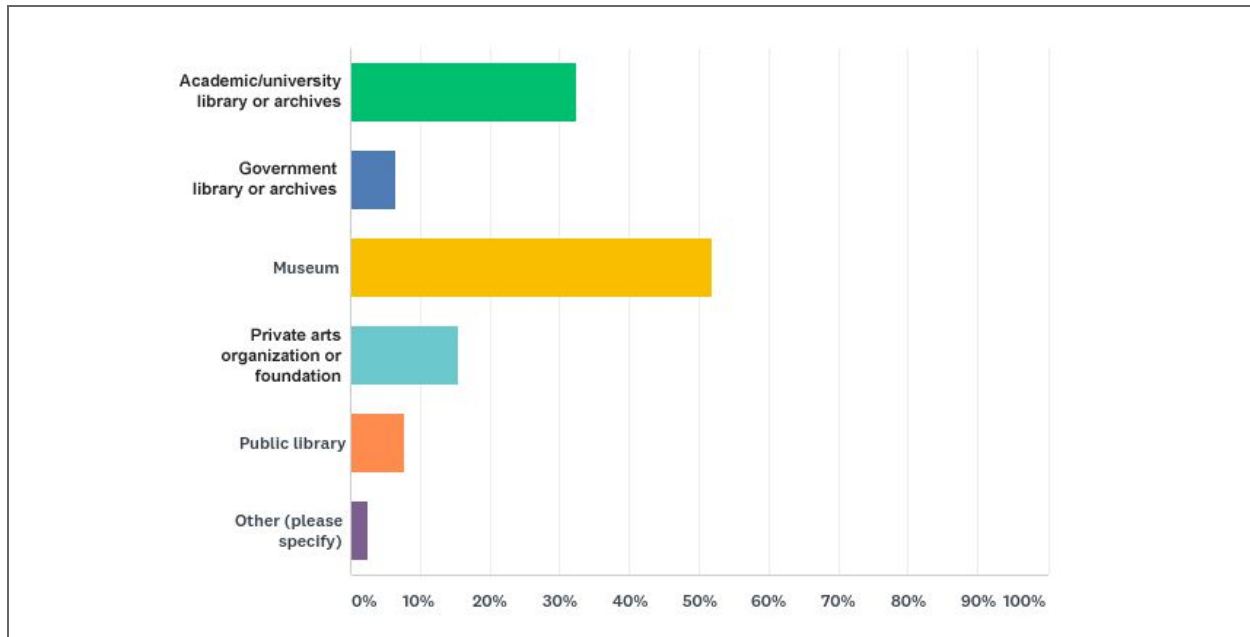
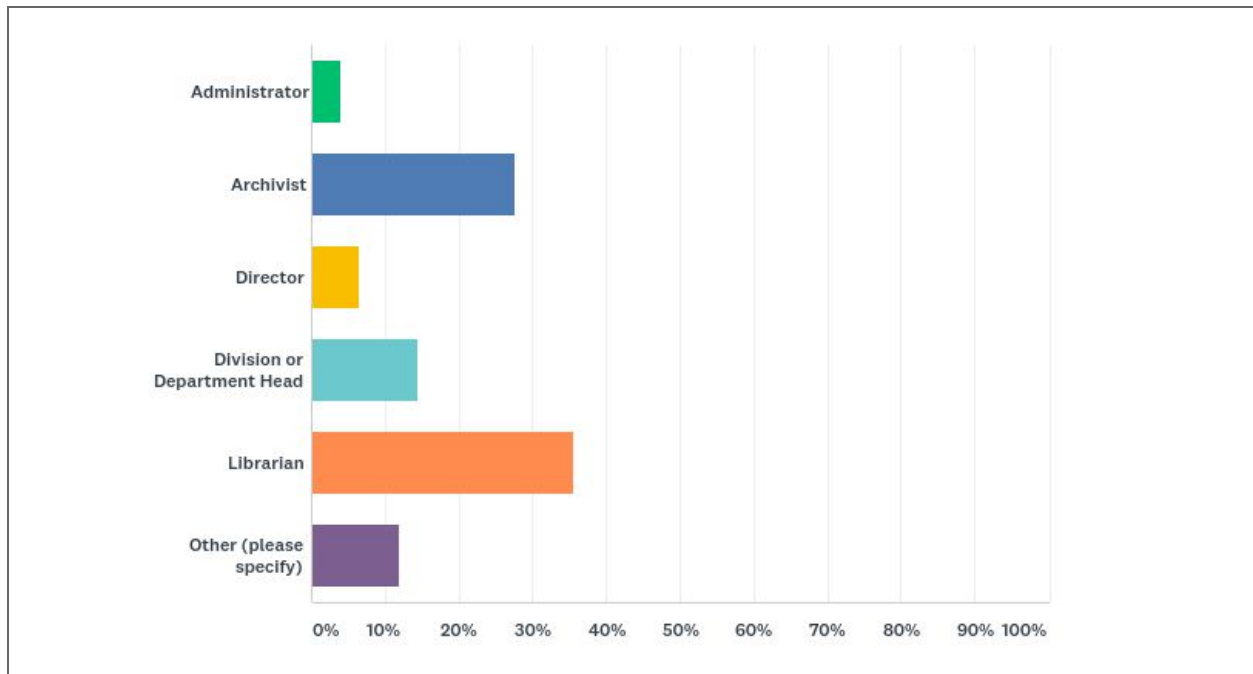


Fig. 1: Respondents' institution type(s)

Respondents were first asked to describe their institutional environment. All respondents (n=77) selected one or more of the provided options and/or provided others. Slightly more than half (52%) identified their institutions as museums and a significant plurality (32%) academic/university libraries or archives. Other responses included a “library within a museum” and college-level academic department. Full results:

Answer choices	Responses	
Academic/university library or archives	32.47%	25
Government library or archives	6.49%	5
Museum	51.95%	40
Private arts organization or foundation	15.58%	12
Public library	7.79%	6
Other (please specify)	2.6%	2

## Respondent role and experience

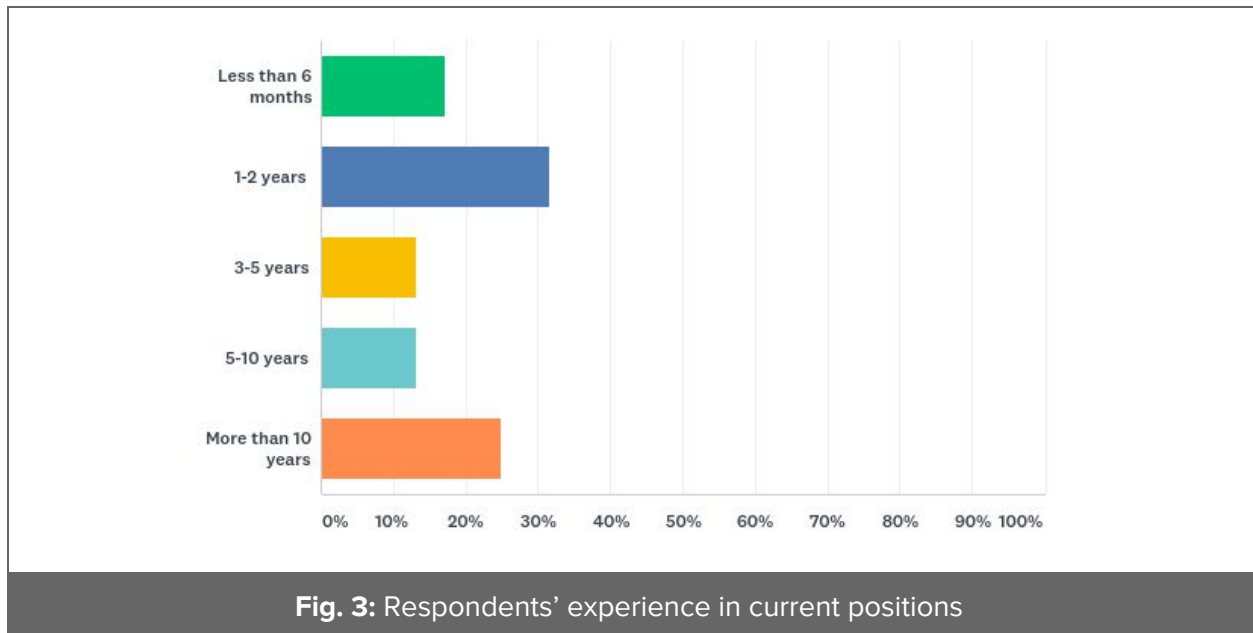


**Fig. 2:** Respondents' professional positions

76 respondents then chose one option each to best describe their own professional role/responsibilities. The largest pluralities were librarians (36%) and archivists (27%). A significant combined population of respondents (21%) also identified as managerial decision makers, "Directors" or "Division or department heads." Other responses included Curator (2), Coordinator, Digital Assets Manager, Digital Contractor, Graduate Student Assistant, Intern/Volunteer, Program Manager, and Project Manager. Full results:

Answer choices	Responses	
Administrator	3.95%	3
Archivist	27.63%	21
Director	6.58%	5
Division or department head	14.47%	11
Librarian	35.53%	27
Other (please specify)	11.84%	9

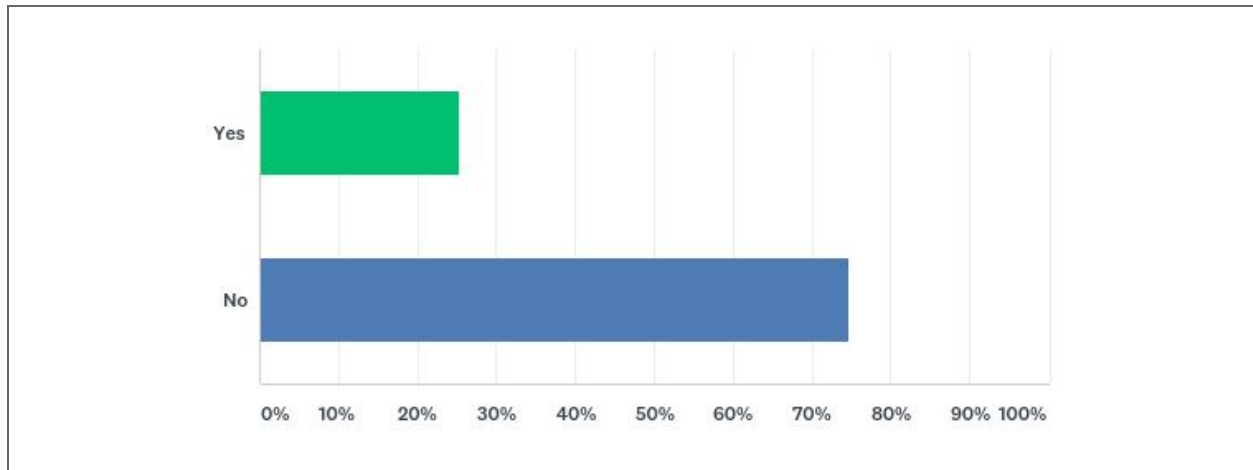
The largest pluralities of respondents' (n=76) experience in their current positions reveal a near-bimodal split between those quite new and most experienced in their respective roles; 32% had been in their current positions for 1-2 years and 25% for more than five years. Responses as a whole were however richly distributed, with no category applying to fewer than 10 respondents, or 13%, each.



Full results:

Answer choices	Responses	
Less than six months	17.11%	13
1-2 years	31.58%	24
3-5 years	13.16%	10
5-10 years	13.16%	10
More than 10 years	25%	19

## Born-digital collecting

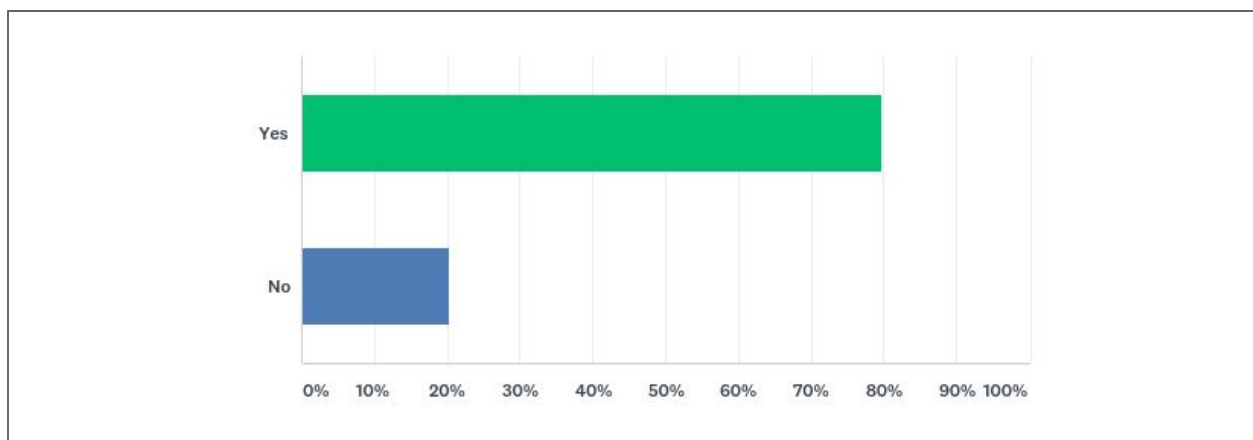


**Fig. 4:** Institutions with born-digital collecting policies

When prompted, the vast majority (75%) of respondents (n=75) indicated that their institutions did not have formal strategies for collecting born-digital materials in general. Full results:

Answer choices	Responses	
Yes	25.33%	19
No	74.67%	56

## Web archiving familiarity and experience

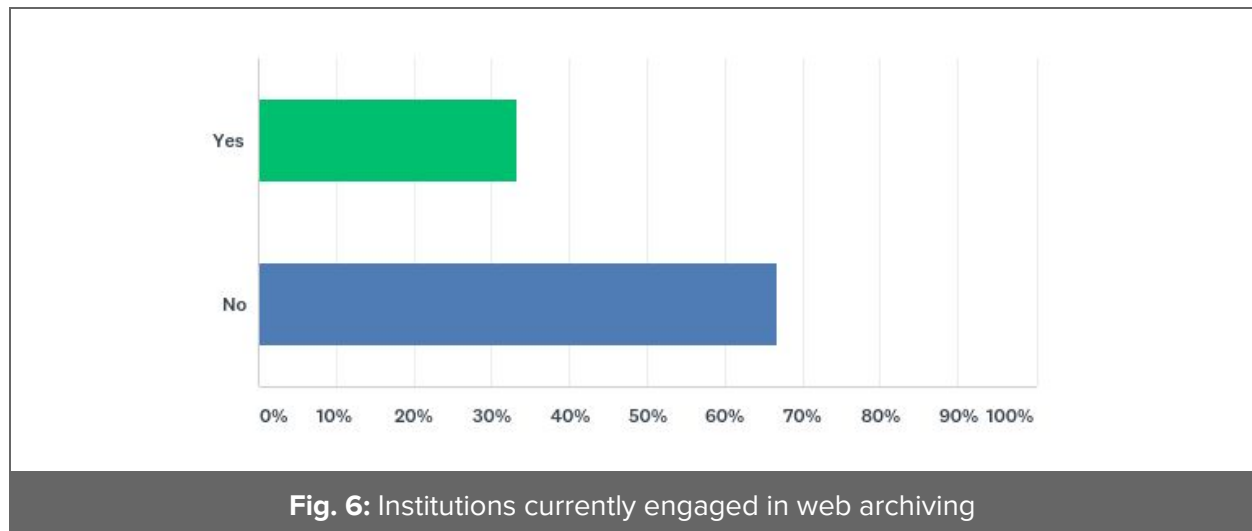


**Fig. 5:** Respondents' familiarity with web archiving as a practice in art and museum libraries

The vast majority (80%) of respondents (n=75) also indicated that they are familiar with web archiving as a practice in art libraries and archives specifically. Full results:

Answer choices	Responses	
Yes	79.73%	59
No	20.27%	15

Respondents who answered “No” to this question were directed next to questions about collecting/acquisitions budgets and descriptive practices towards the survey’s end. Those who responded “Yes” to or skipped this question proceeded to the next question. The clear majority (67%) of the latter group (n=60) indicated that their institutions do not currently engage in the specific practice of web archiving.

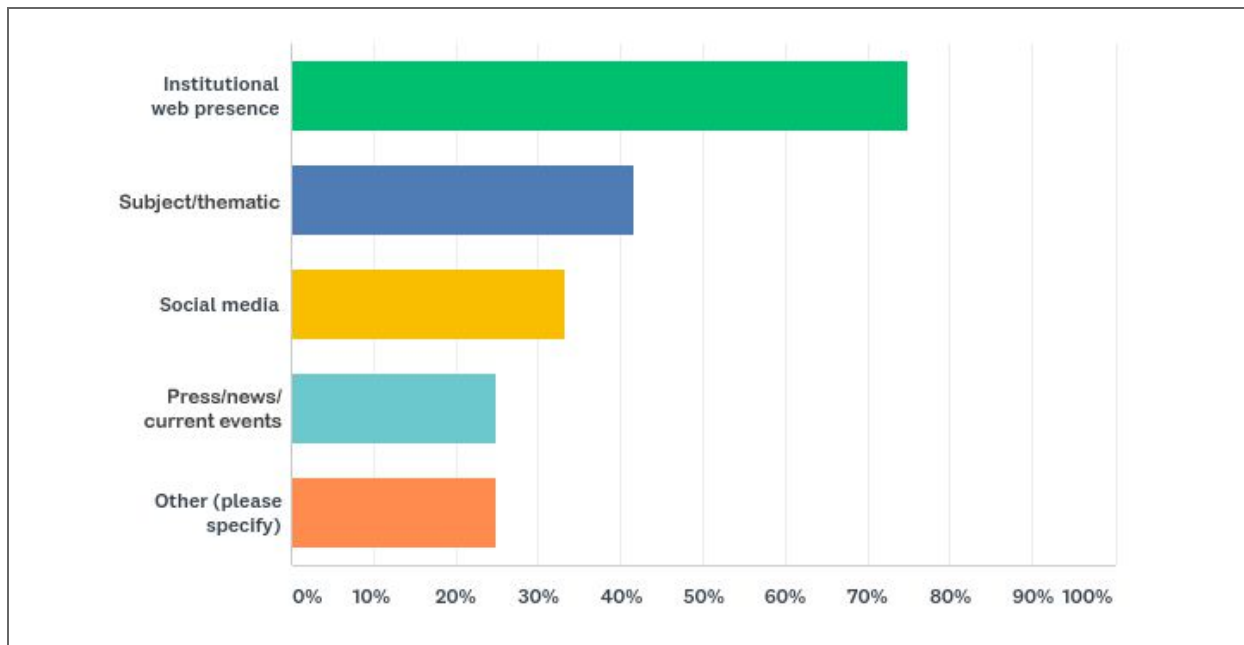


Full results:

Answer choices	Responses	
Yes	33.33%	20
No	66.67%	40

Respondents who indicated that their institutions do not currently engage in web archiving were subsequently directed beyond the four following questions about those web archiving programs.

## Current practices and interests

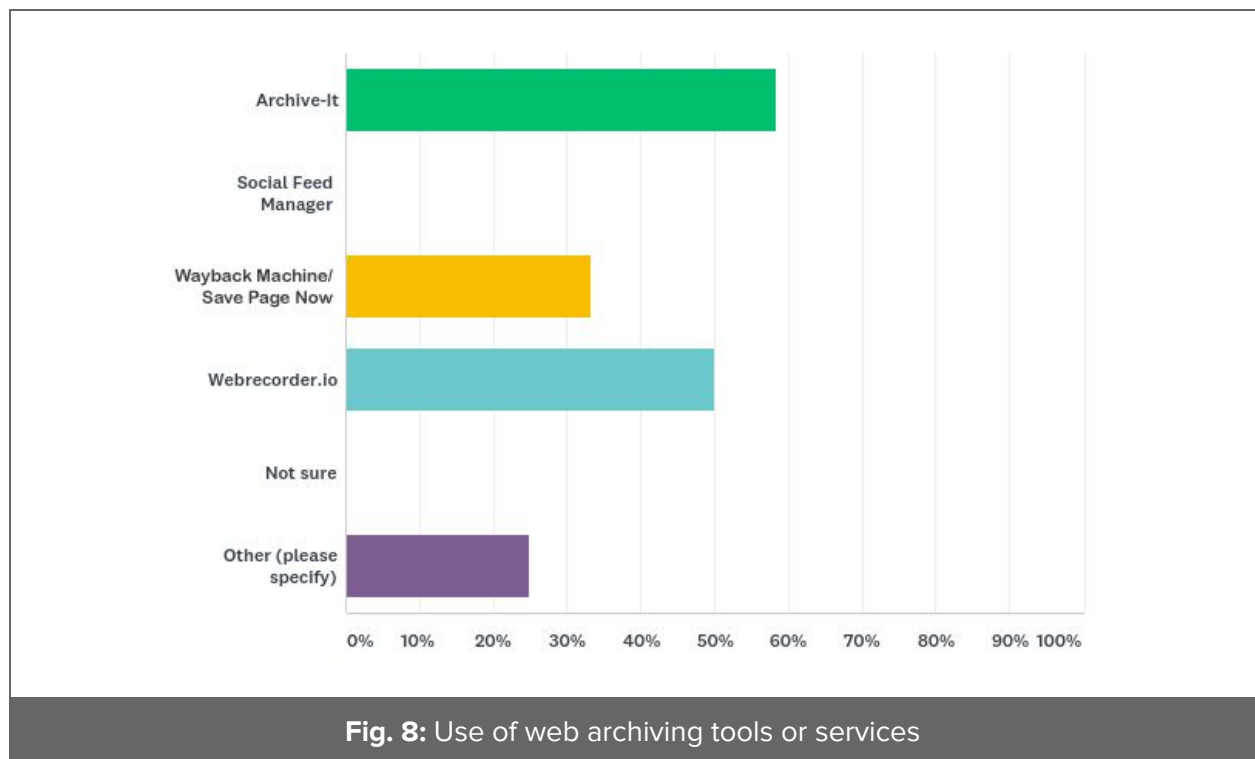


**Fig. 7:** Current web archive collecting areas

Respondents who indicated that their institutions were currently engaged in web archiving (n=20) were asked to describe their web collecting scopes. Responses to this question (n=12) indicate a current emphasis on these institutions' own web presences, while smaller but significant groups also collect thematically and/or specifically from third party social media or press platforms. Other responses to this question included "project websites," student and faculty projects, and "pilot captures of 2 arts organizations." Full results:

Answer choices	Responses	
Institutional web presence	75%	9
Subject/thematic	41.67%	5
Social media	33.33%	4
Press/news/current events	25%	3
Other (please specify)	25%	3

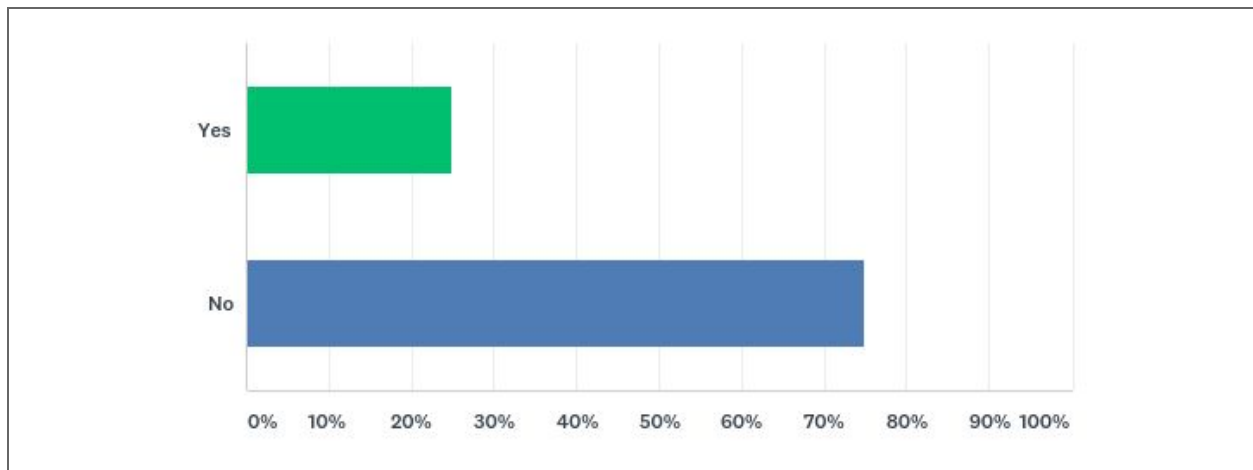
Question eight asked this same survey sub-group to provide the project team with direct links to information about their extant web archives. Responses (n=6) included identifying information about the respondents and are not published here.



These respondents were further asked to identify what if any web archiving tools or services their institutions utilize. A slight majority (58%) of total responses (n=12) identified the Internet Archive's Archive-It service, half Rhizome's Webrecorder, and one-third also indicated that they use the "Save Page Now" tool that adds individual web pages to the Internet Archive's Wayback Machine upon user command. Other responses included: WinHTTP, "Heritrix, PDF," and an institution that has not adopted any specific tools for a program that is currently in its planning phase. Full results:

Answer choices	Responses	
Archive-It	58.33%	7
Social Feed Manager	0%	0
Wayback Machine/Save Page Now	33.33%	4
Webrecorder.io	50%	6
Not sure	0%	0
Other (please specify)	25%	3

The twelve respondents in the sub-group of institutions that currently have programs were further asked if their institutions have staff devoted to web archiving.

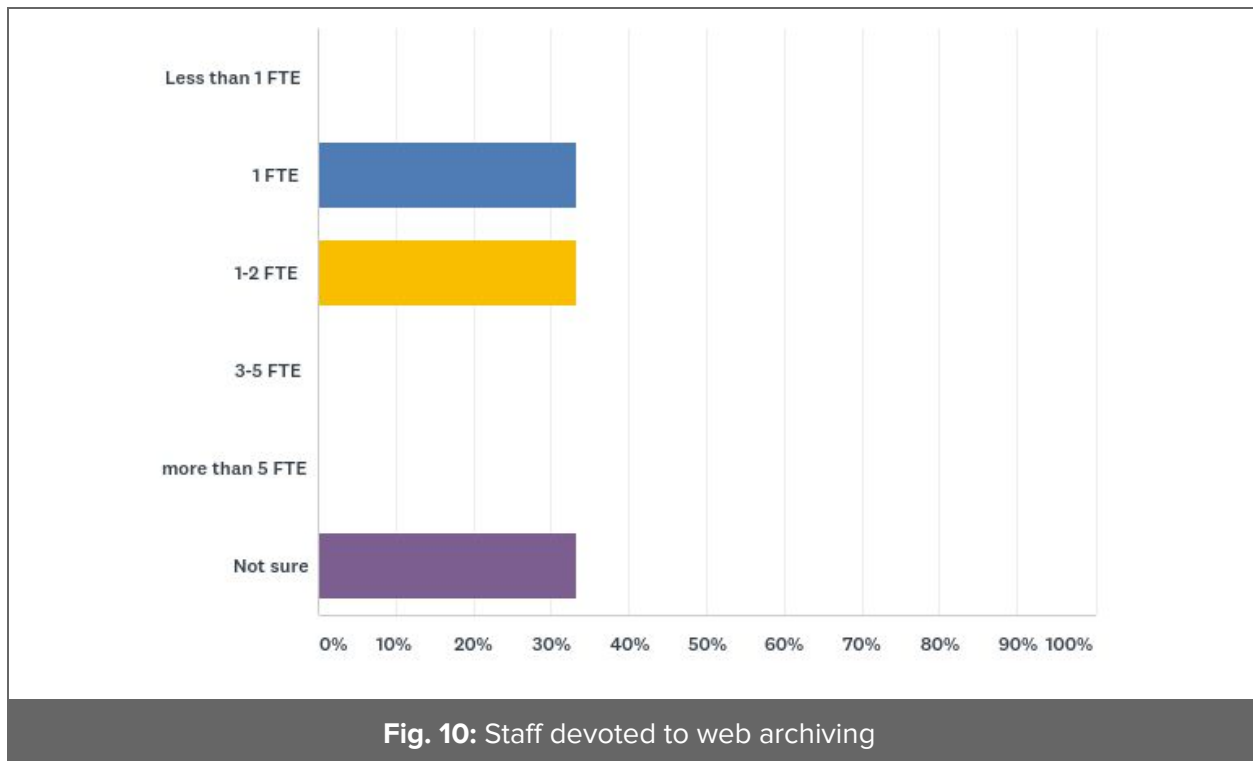


**Fig. 9:** Use of web archiving tools or services

The vast majority (75%) responded that they do not. Full Results:

Answer choices	Responses	
Yes	25%	3
No	75%	9

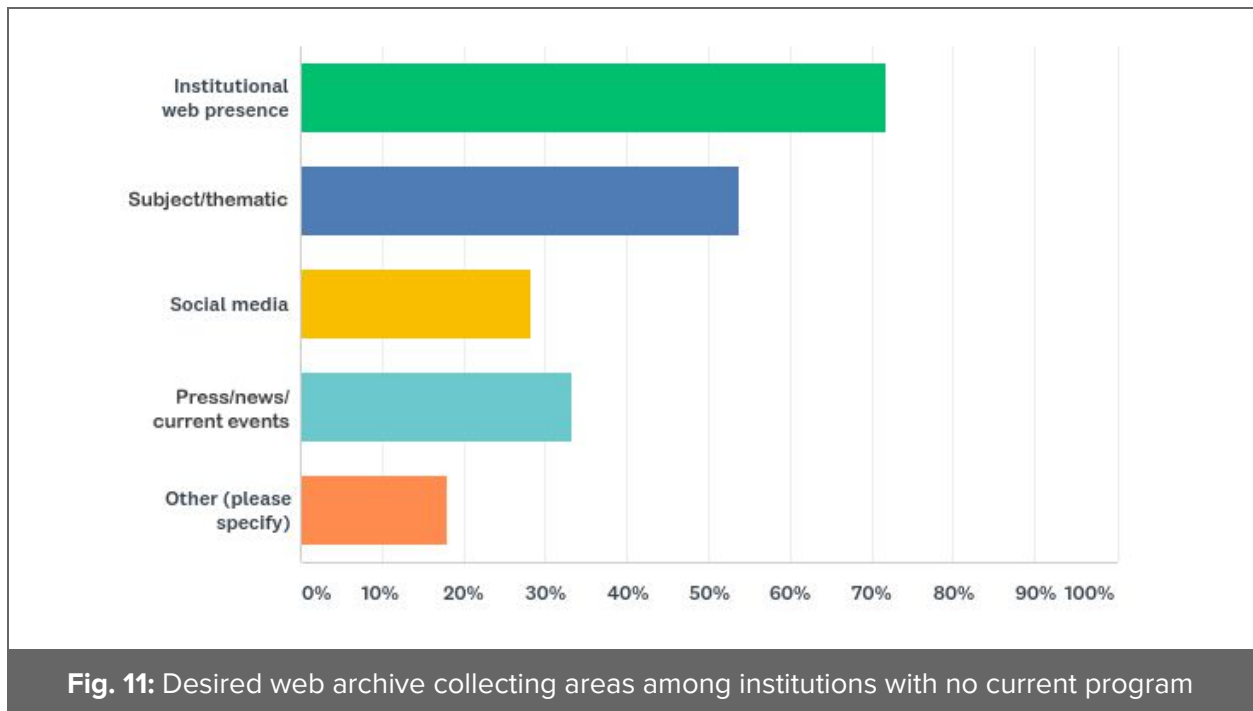
Respondents who responded “No” to this question were directed beyond the following two questions about staffing. The three respondents who responded that they did indeed have staff devoted to web archiving were further asked how many staff that their institutions employed to this end.



One responded that 1 full-time equivalent (FTE) employee is devoted to web archiving, one that 1-2 FTE are, and a third that they were not sure. Full results:

Answer choices	Responses	
Less than 1 FTE	0%	0
1 FTE	33.33%	1
1-2 FTE	33.33%	1
3-5 FTE	0%	0
More than 5 FTE	0%	0
Not sure	33.33%	1

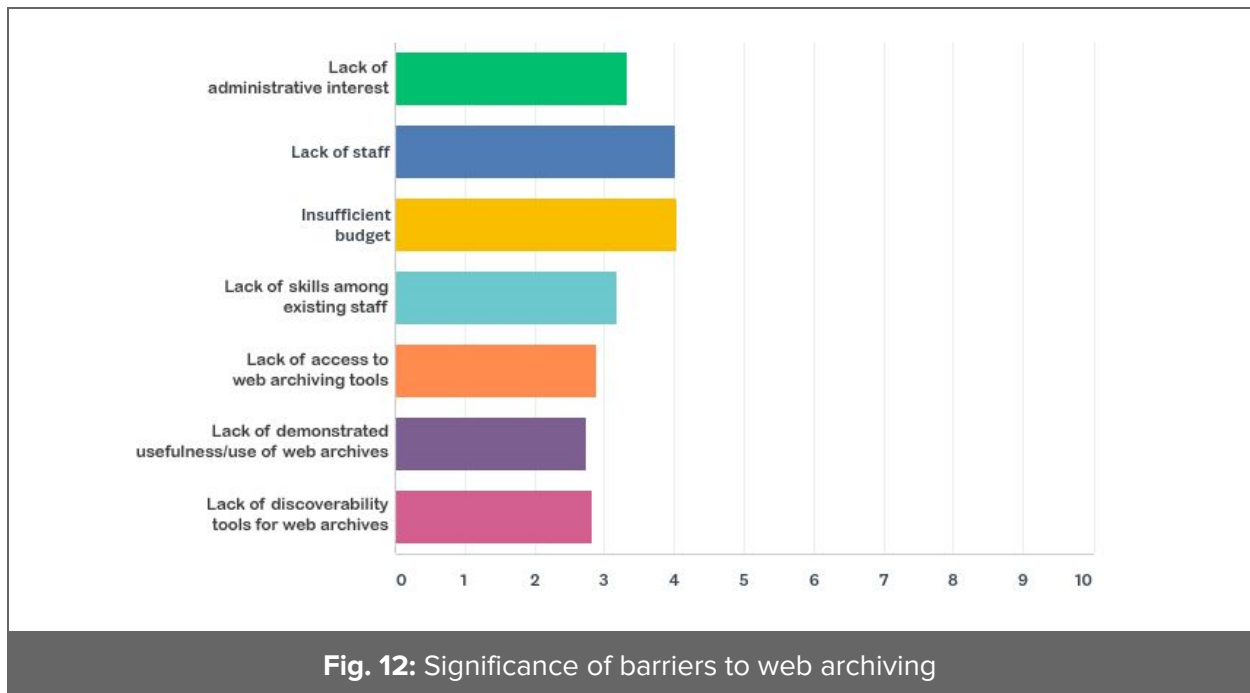
Respondents who indicated that their institutions did *not* currently have one were in turn asked to indicate the areas of focus that they would pursue with a web archiving program.



Respondents to this question (n=39) likewise prioritized institutional records on the web highest (72%), but a majority also expressed their interest in collecting thematically (54%). Moreover, an additional five respondents volunteered other areas that the project team considers included under the category of subject/thematic, including artist files, auction catalogues, and born-digital exhibitions. One other respondent was “not sure” and one respondent volunteered “Historic Area Photograph Collection,” perhaps misunderstanding the web archiving context of the question. Full Results:

Answer choices	Responses	
Institutional web presence	71.79%	28
Subject/thematic	53.85%	21
Social media	28.21%	11
Press/news/current events	33.33%	13
Other (please specify)	17.95%	7

## Barriers to web archiving

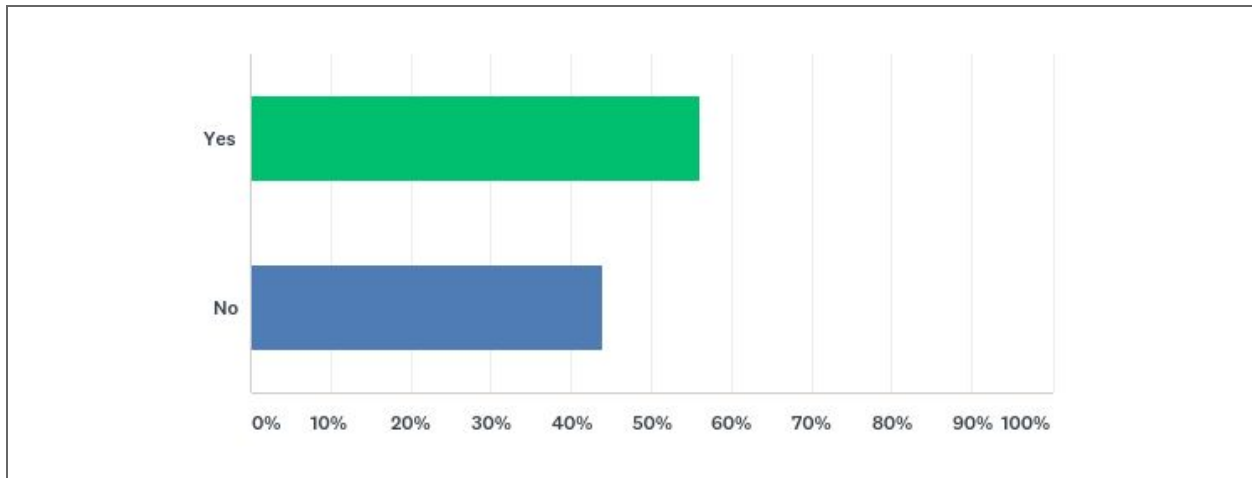


All survey participants who responded that they were familiar with web archiving in art and museum libraries or skipped this question were asked to rate the significance of specified barriers to web archiving at their institutions on a traditional Likert scale from 1 (low barrier) to 5 (high barrier). Responses (n=50) identified budget and staffing issues highest. These were the only two response choices to transcend a weighted average of 4. Comparatively lower barriers--those below a weighted average response of 3--included perceived lacks of access to web archiving tools, discovery tools, and demonstrated use cases. Full results:

Answer choices	Weighted average
Lack of administrative interest	3.33
Lack of staff	4.02
Insufficient budget	4.04
Lack of skills among existing staff	3.17
Lack of access to web archiving tools	2.89
Lack of demonstrated usefulness/use of web archives	2.74
Lack of discoverability tools for web archives	2.82

Free-form comments invited in response to this question ranged from identification of additional challenges--such as administrative support for outreach or IT department support for technical requirements--to comments on the prioritization of web archiving among other pressing needs and challenges, suggesting for instance that it is “perpetually back-burnered.”

## Web archiving training

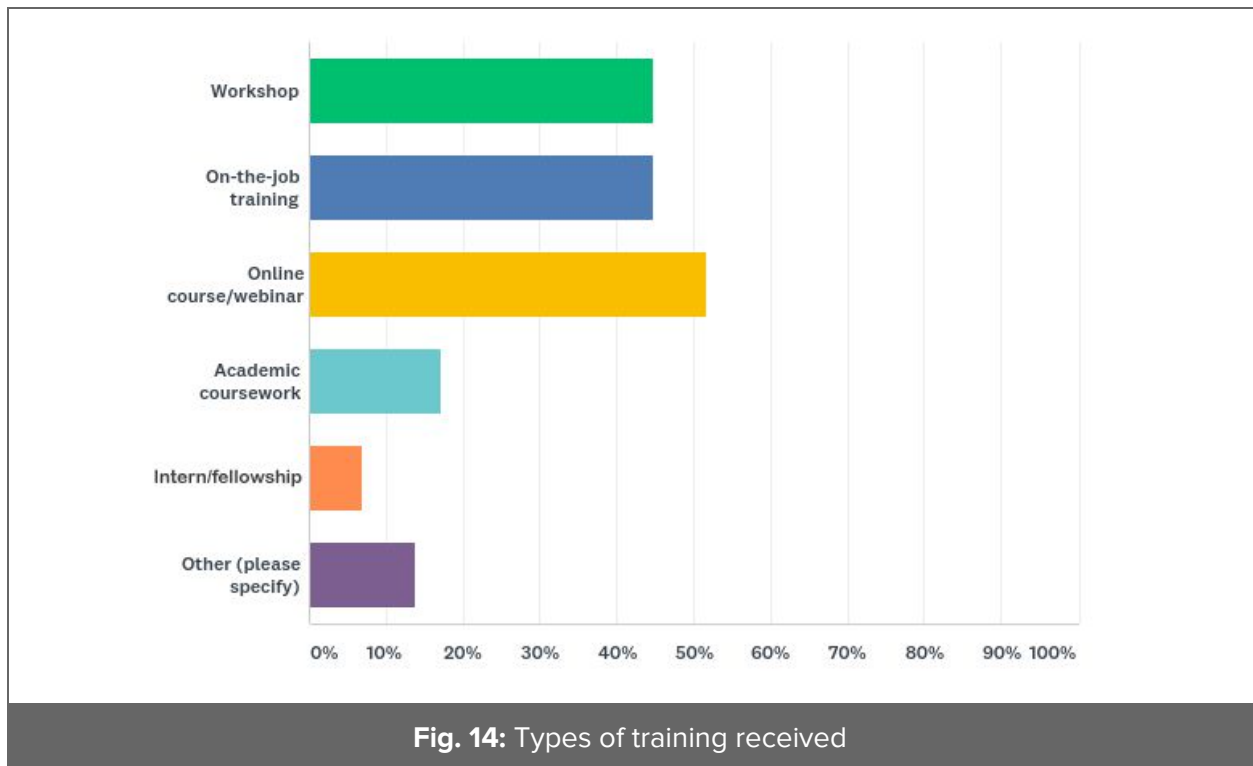


**Fig. 13:** Respondents’ training in web archiving

When then asked if they have received any training in web archiving, a majority (56%) of respondents (n=50) indicated that they had. Full results:

Answer choices	Responses	
Yes	56%	28
No	44%	22

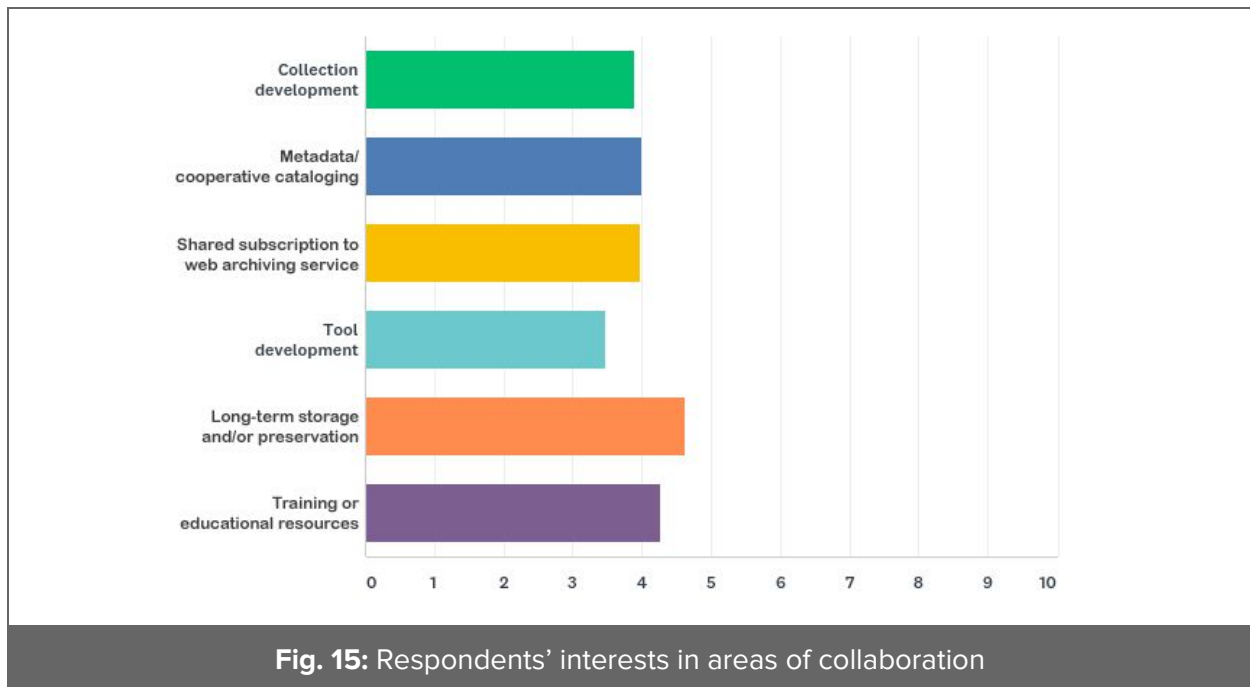
When subsequently asked to specify the forms of training received, a slight majority (52%) of respondents (n=29) indicated attending an online course/webinar, while significant pluralities (45% each) also identified workshops or on-the-job training.



Other responses volunteered by respondents included: conference presentations, “hands on work with Archive-It account,” and published literature. Full results:

Answer choices	Responses	
Workshop	44.83%	13
On-the-job training	44.83%	13
Online course/webinar	51.72%	15
Academic coursework	17.24%	5
Intern/fellowship	6.9%	2
Other (please specify)	13.79%	4

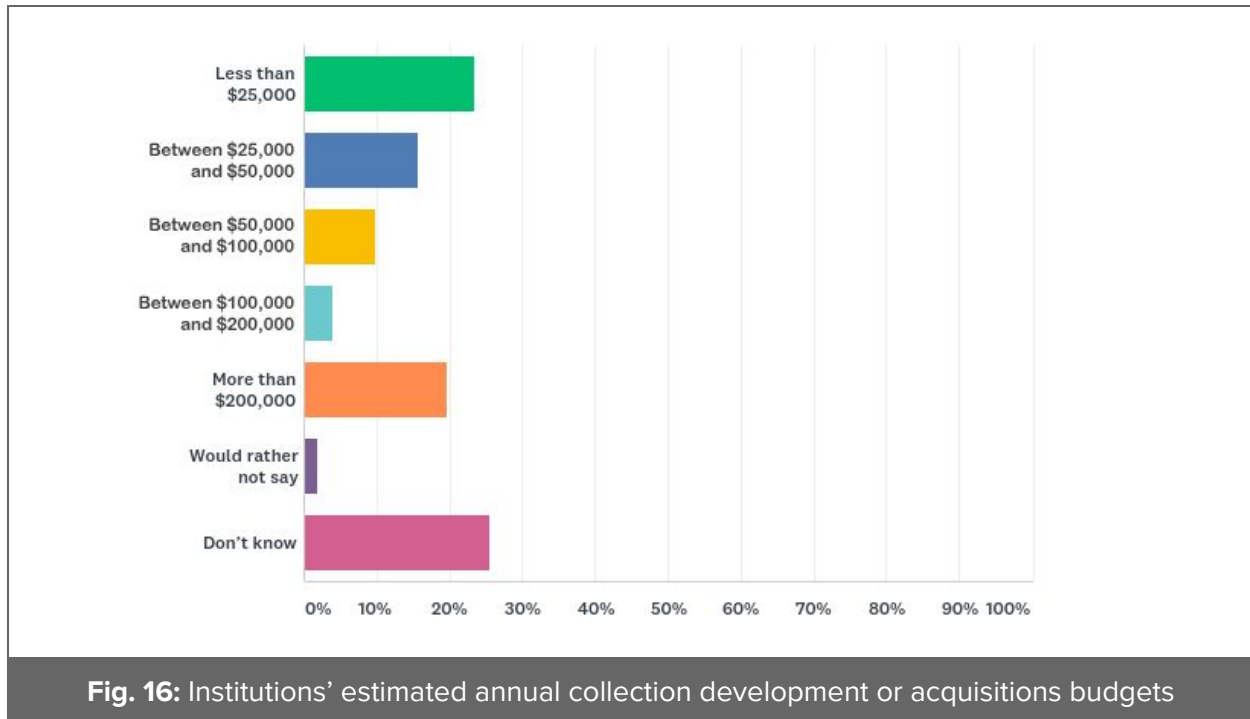
## Interests in collaboration



Respondents who did not specifically volunteer that they were *unfamiliar* with the practice in Question 5 were asked to rate their interests in collaboration opportunities for web archiving among art and museum libraries. Respondents (n=50) rated options provided by the project team, again on the Likert scale from 1 (not interested) to 5 (very interested). Their responses indicate support for all options provided, which each exceeding a weighted average of 3, but especially significant support for collaboration on issues of long-term storage and/or preservation. Full results:

Answer choices	Weighted average
Collection development	3.9
Metadata/cooperative cataloging	4
Shared subscription to web archiving service	3.98
Tool development	3.47
Long-term storage and/or preservation	4.63
Training or educational resources	4.26

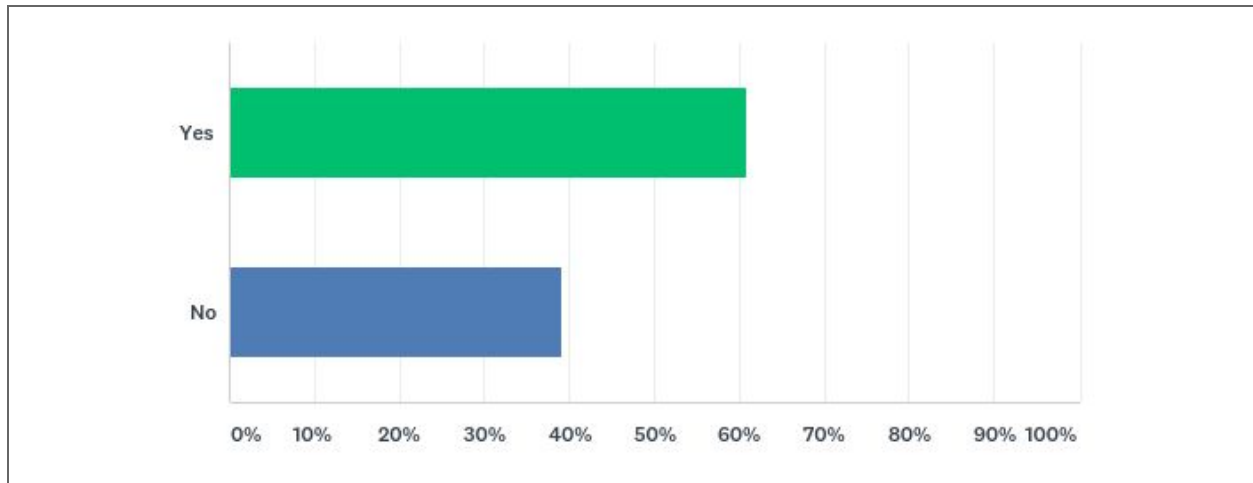
## Budgets



All survey participants were asked to classify their scale of budget for collections/acquisitions. The largest plurality (25%) among all respondents (n=51) did not know how to answer this question. The next largest groups however suggest a near-bimodal split again among respondents at the higher and lower ends of the spectrum. Full results:

Answer choices	Responses	
Less than \$25,000	23.53%	12
Between \$25,000 and \$50,000	15.69%	8
Between \$50,000 and \$100,000	9.8%	5
Between \$100,000 and \$200,000	3.92%	2
More than \$200,000	19.61%	10
Would rather not say	1.96%	1
Don't know	25.49%	13

## Descriptive practice

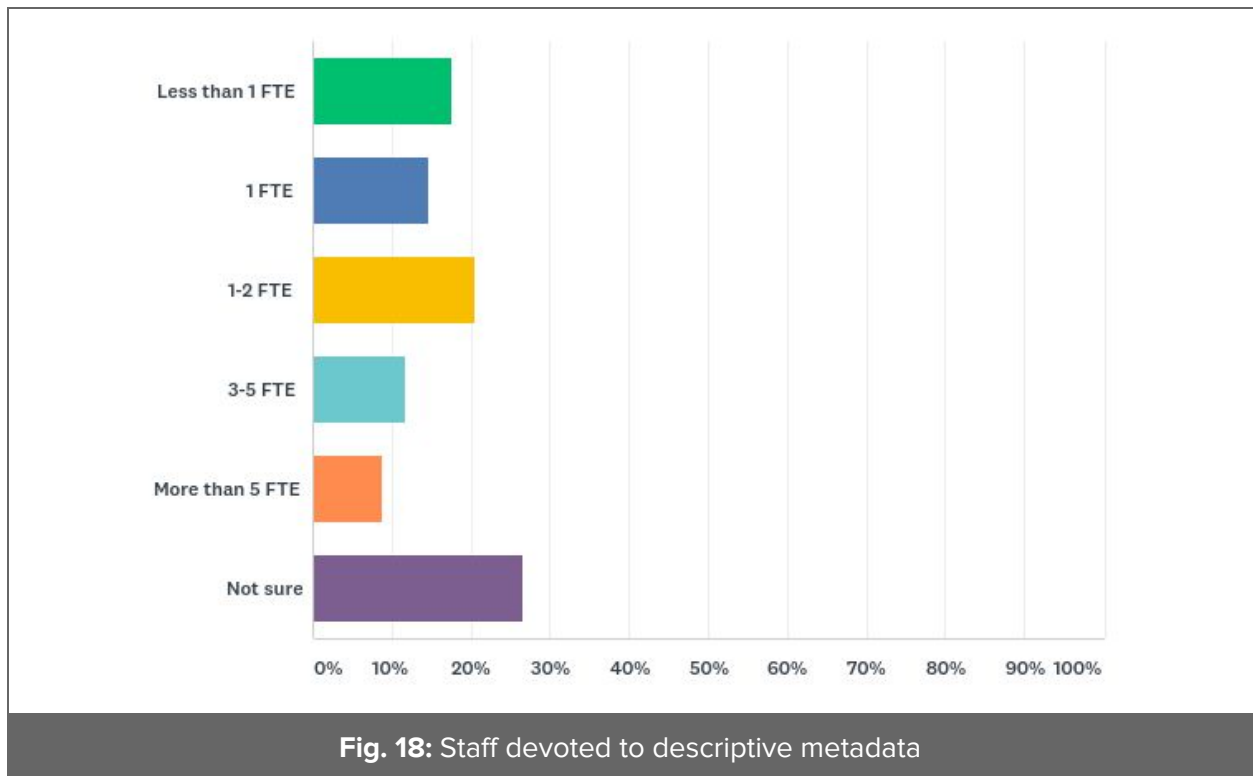


**Fig. 17:** Institutions with staff devoted to descriptive metadata

All participants were asked to report if their institutions devoted to staff to descriptive metadata. The clear majority (61%) of respondents (n=51) indicated that they do. Full results:

Answer choices	Responses	
Yes	60.78%	31
No	39.22%	20

Those survey participants who answered “Yes” to this question were asked to report how many FTE staff were devoted to description at their institutions.



Again the largest plurality (26%) among all respondents (9) reported that they did not know this. Other responses indicated a diverse range of staff volume devoted to this area across art among respondents' institutions, from less than 1 FTE up to more than 5 FTE. Full results:

Answer choices	Responses	
Less than 1 FTE	17.65%	6
1 FTE	14.71%	5
1-2 FTE	20.59%	7
3-5 FTE	11.76%	4
More than 5 FTE	8.82%	3
Not sure	26.47%	9

Remaining survey questions asked survey participants for direct comments on the survey and larger project, and for institutional and contact data. Results were identifying and not published here.

## ANALYSIS

This survey engaged a much broader community of art and museum library professionals than had prior research into web archiving programs and their uses, notably the *Web Archiving in the United States* survey reports published by the National Digital Stewardship Alliance (Farrell et al, 2018) and *Web Archiving Environmental Scan* published by Harvard Library (Truman, 2016), in which only a select few participated. The results above help to paint a richer, more representative picture of the community, it's interests, and challenges, about which new conclusions may be drawn:

### Community

**Art and museum libraries are familiar with web archives, aware of their values, and have practical concerns about creating and hosting them.**

The project team's email invitations stressed the value of responses even from professionals with no prior exposure to web archives, 15 of whom responded and provided other useful information about their institutions and roles. That 80% of respondents nonetheless affirmed their familiarity with web archiving as a practice in art and museum libraries, 56% reported that they have even received some training, and 33% responded from institutions that currently engage in some form of web archiving, underlines that the basic values and uses of web archives are broadly known and understood in this community.

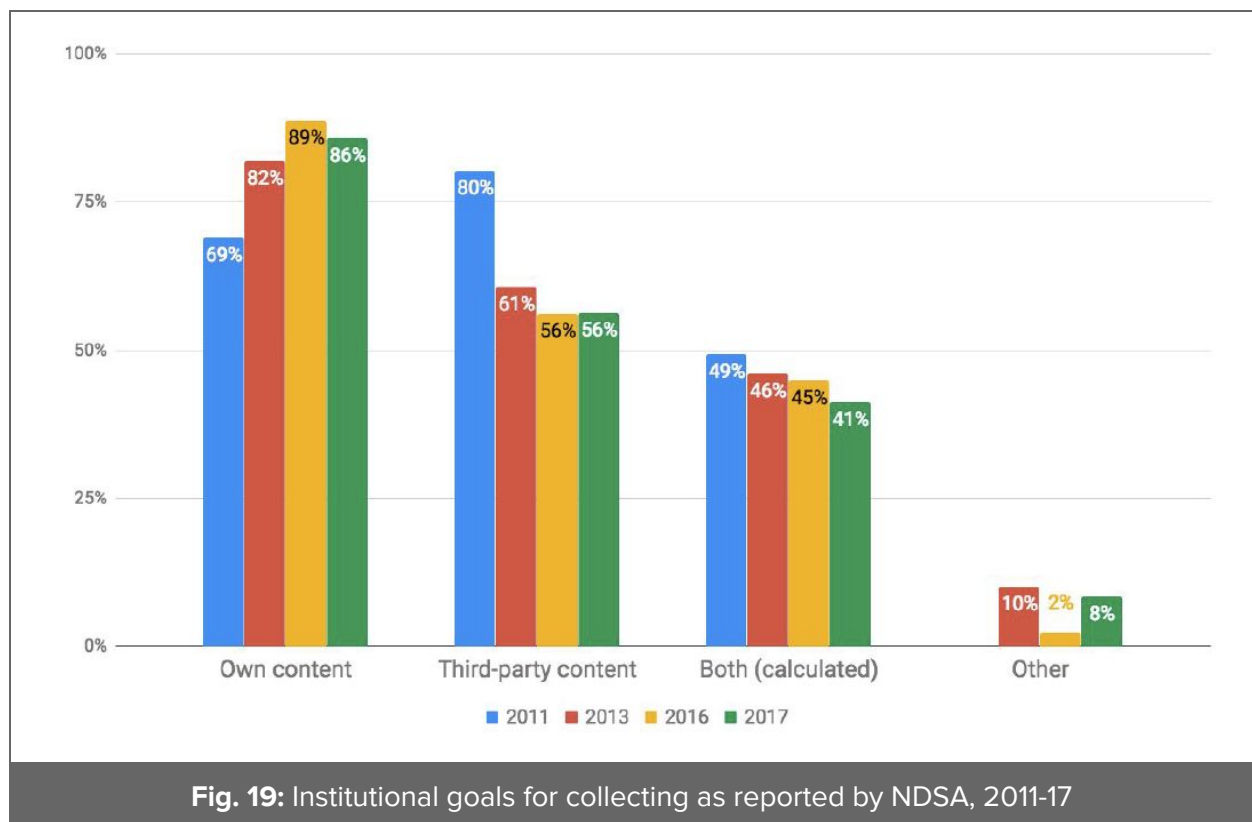
These results cut across distinctive intersections of respondent types. 24 respondents with five years' or more experience in their positions reported that they are familiar with web archiving as a practice in art and museum libraries, compared to only three who are not. 13 who occupy managerial decision-making roles--Directors or Department/Division Heads--reported that they are likewise familiar, compared to only three who are not. Six furthermore reported that they have trained in web archiving, compared to five who reported that they had not. 16 respondents with two or fewer years' experience in their positions reported that they have trained, compared to only four who have not. Web archiving is understood at senior and influential levels in art museum libraries, whose workforces are increasingly prepared to implement the programs.

Accordingly, the community's engagement with web archives is more obstructed by practical, material issues affecting implementation than by demonstrated values and use cases. They in fact ranked "Lack of demonstrated usefulness/use of web archives" as the least significant barrier to web archiving at their own institutions. As discussed further in the *Obstacles and opportunities* section below, accessibility to the financial and human resources necessary to get started is a far more important impediment, and the opportunity to share the long-term infrastructural resources to maintain them is the most common goal for collaboration.

## Collecting

**Art and museum libraries want to collect, manage, and preserve a wide and varied corpus of web-based material.**

75% of institutions with active programs archive their own web presences, and 80% of those without them would do the same. A significant plurality of the former group (42%) also collect to complement or enhance the subjects and themes in their reference and research collections, and majority of the latter (54%) expressed their interest in doing the same. The regular NDSA surveys of the wider web archiving community record a trend towards a similar ratio of content collecting goals, with 86% of respondents reporting in 2017 that they archive their own institutions' web content and 56% that they archive third-party content.



This survey's respondents were also asked to record their interests in other specific subjects and forms that may cut across the two larger categories of 'internal' and 'external' content, and which in turn registered lower levels of interest each. It is however notable that no provided option was of interest to less than 25% of respondents with active web archiving operations nor to 28% of those without. Moreover, the striking majority of respondents to this question (28 of 39, or 72%) identified more than one area of interest.

No one software or service quite dominates yet in supporting this goal. Among the twelve respondents who reported using specific tools or software to collect from the web, fully half (6) use more than one of the provided options. No follow-up questions were asked about these options, so engaging representative web archiving programs to speak to the advantages and limitations of each would help technologists to address any perceived gaps on their development roadmaps and/or collaborate more directly on meeting the needs of art and museum libraries among them.

## Obstacles and opportunities

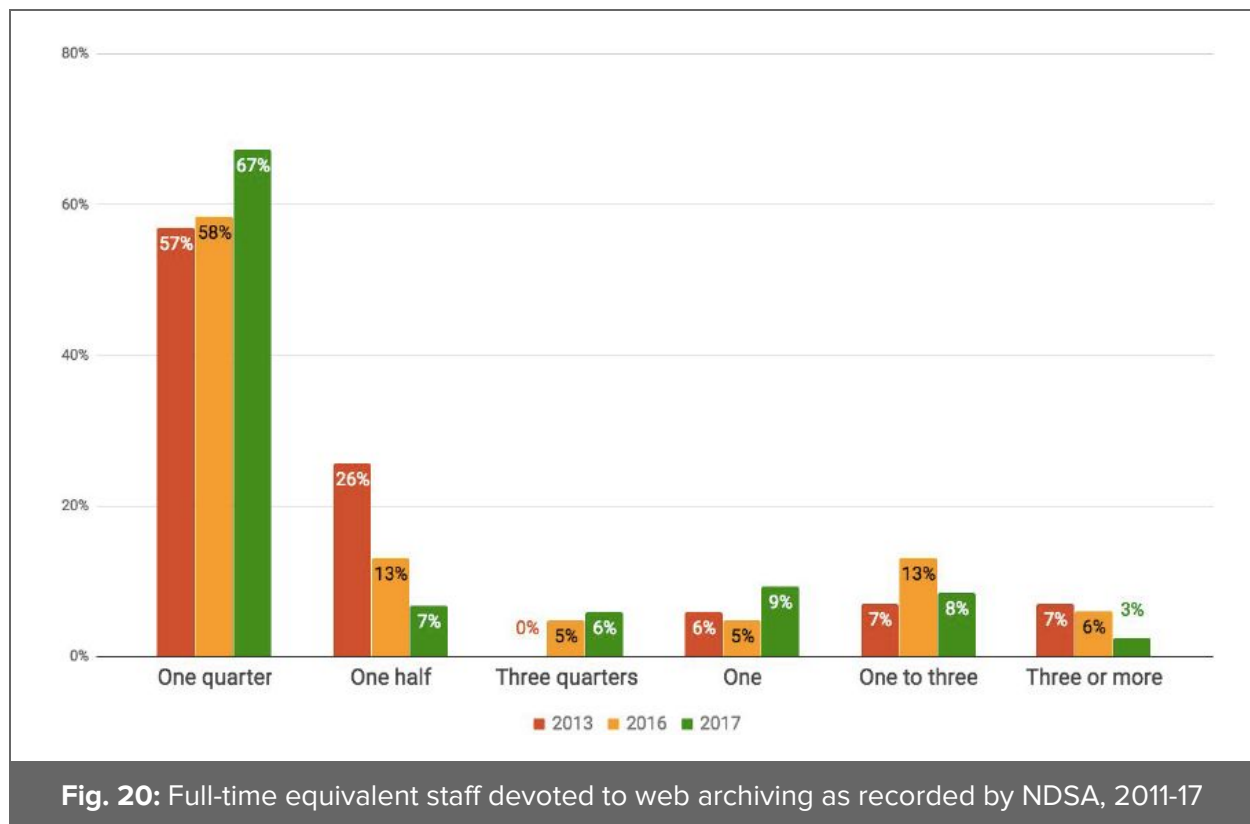
**Across art and museum library types and perspectives, the most pressing needs and opportunities for progress surround appropriating the human and material resources to both engage and sustain web archives for the longest term.**

Survey respondents expressed several barriers to web archiving, however none so consistently as insufficient budget or lack of staff. These material support factors were the closely-weighted top obstacles to all types and intersections of respondents--regardless of institutional typology, professional responsibility or seniority, collecting or acquisitions budget. These were even the most significant barriers reported by respondents from institutions currently engaged in web archiving, though in this rare case the lack of staff was weighted as even more significant than the insufficient budget.

The amount of staffing necessary to maintain or participate in a web archive is however still an open question for this community. The project team believe that there may have been ambiguity and misunderstanding surrounding these survey questions. While 20 respondents reported that their institutions are currently engaged in web archiving, only three also reported that their institutions devote staff to the effort, and none that any fewer than 1 FTE were devoted specifically. This suggests that the initial question precluded the idea that devoting less than one staff member's full portfolio to web archives would be an option. As commented by one respondent later:

on previous question, a staff member is not devoted to it but as the archivist, a portion of my time is spent capturing our institutional web presence--probably 3 hours/month at this point.

The NDSA web archiving surveys have however consistently found that devoting  $\leq 1$  FTE--and increasingly even  $\leq 0.25$  FTE--is by far the most common practice. Art and museum libraries with current programs should be empowered to report out more fully on the staffing allocations made to build and sustain their web archives.



Respondents were also remarkably consistent across types and intersections in weighting their priorities for collaboration with their peers in art and museum libraries. All provided options were deemed sufficiently significant by the community to probe further at and beyond the National Forum events, however none as consistently as the issue of long-term storage and preservation of their web archives. Web archiving technologists allied professionals in academic libraries and archives can speak to their precedent efforts in these areas, bringing together stakeholders and laying this foundation of trust for higher levels of collaboration.

Strategic coordination is in fact a more determinative indicator of art and museum libraries' engagement in web archiving than any other found in this survey. This includes the sheer volume of institutions' collecting or acquisitions budgets. Of the institutions at the highest end of the budget scale (More than \$200,000), only two engage actively in web archiving and only one has a more general strategy for collecting born-digital media. They trail behind their lesser-resourced peers in both areas. Among institutions that do have born-digital collecting strategies however, seven reported that they collect from the web, compared to eight who do not. While still a slight minority, this group of strategically directed collectors outpaces the one-third of the broader community participating in web archiving. And while this group's most pained points are also funding and staffing, their experience and perspectives on strategically allocating sufficient resources to sustain web archiving at their institutions would be especially valuable.

## REFERENCES

- Farrell, M., McCain, E., Praetzellis, M., Thomas, G., & Walker, P. (2018). *Web archiving in the United States: A 2017 survey*. National Digital Stewardship Alliance.  
doi:10.17605/OSF.IO/3QH6N4
- Internet Archive & New York Art Resources Consortium. (2018). *Advancing art libraries and curated web archives: A literature review*.  
<https://archive.org/details/AALCWALiteratureReview>
- Truman, G. (2016). *Web archiving environmental scan: Harvard Library report*.  
<http://nrs.harvard.edu/urn-3:HUL.InstRepos:25658314>